

企业级Hadoop平台源码二次开发-第二天

一、课前准备

掌握Hadoop RPC基本原理

掌握HDFS 高可用原理

二、课堂主题

通过讲解HDFS的元数据流程，让学生熟悉分布式技术的本质，为后面的二次开发和学习其他大数据技术打下基础

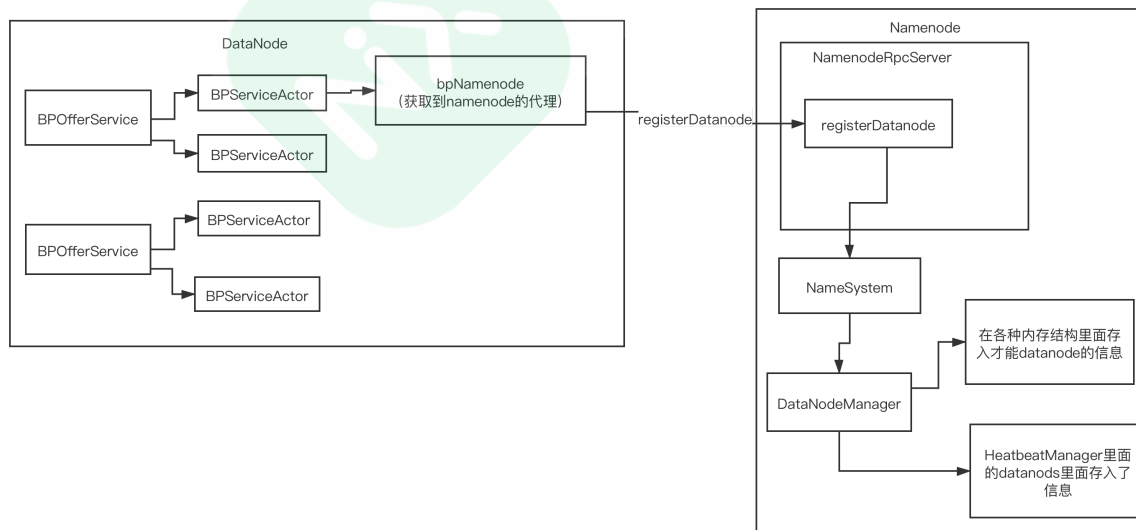
三、课程目标

1. 掌握DataNode启动流程
2. 掌握NameNode管理元数据的流程
3. 掌握HDFS源码中的优秀架构思想
4. 掌握HDFS源码中的优秀代码设计

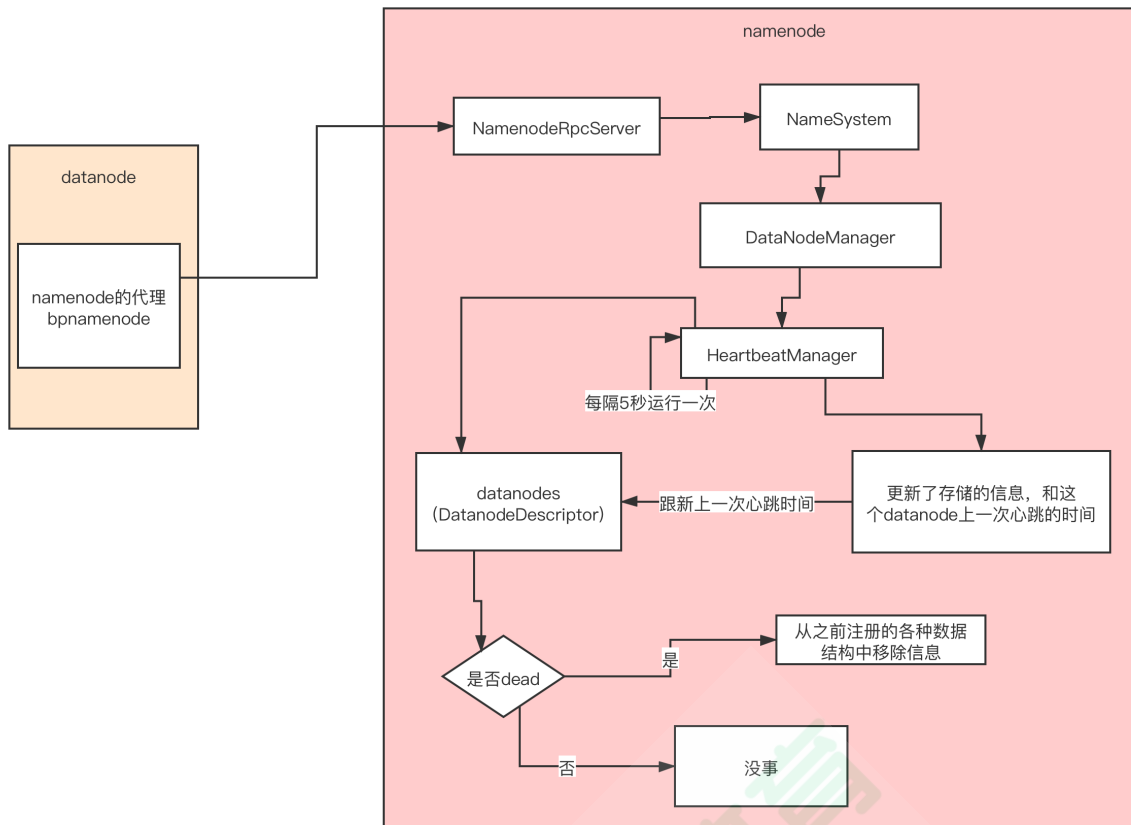
四、知识要点

4.1 DataNode注册 (30分钟)

4.1.1 DataNode注册

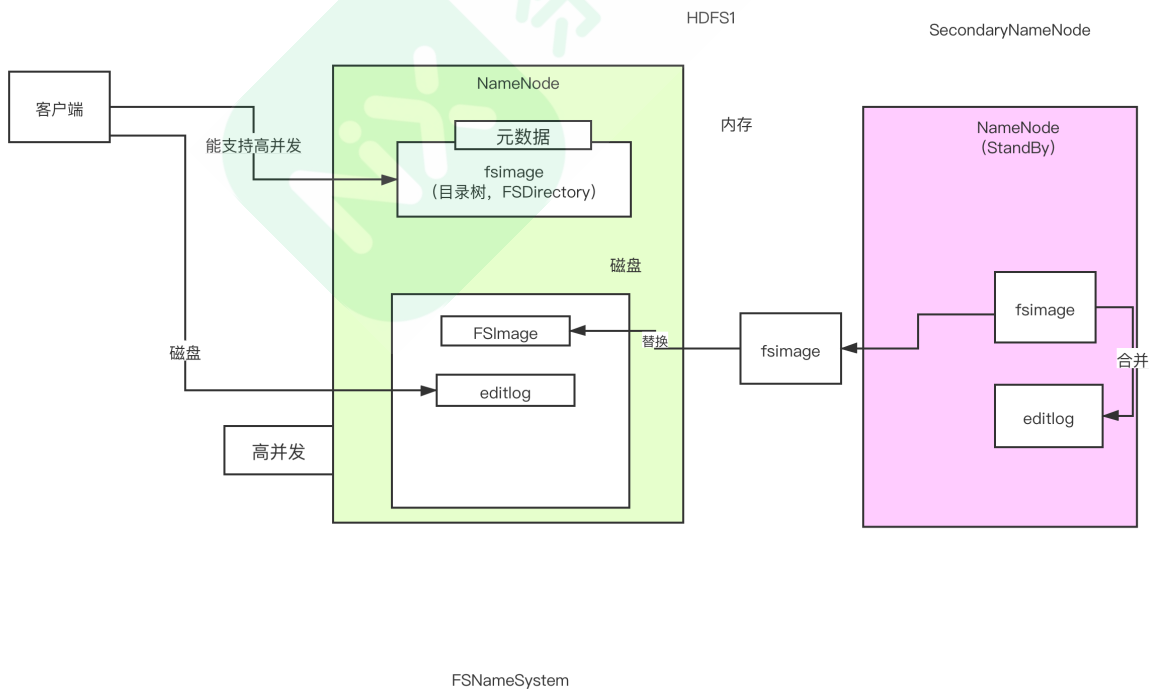


4.1.2 DataNode心跳

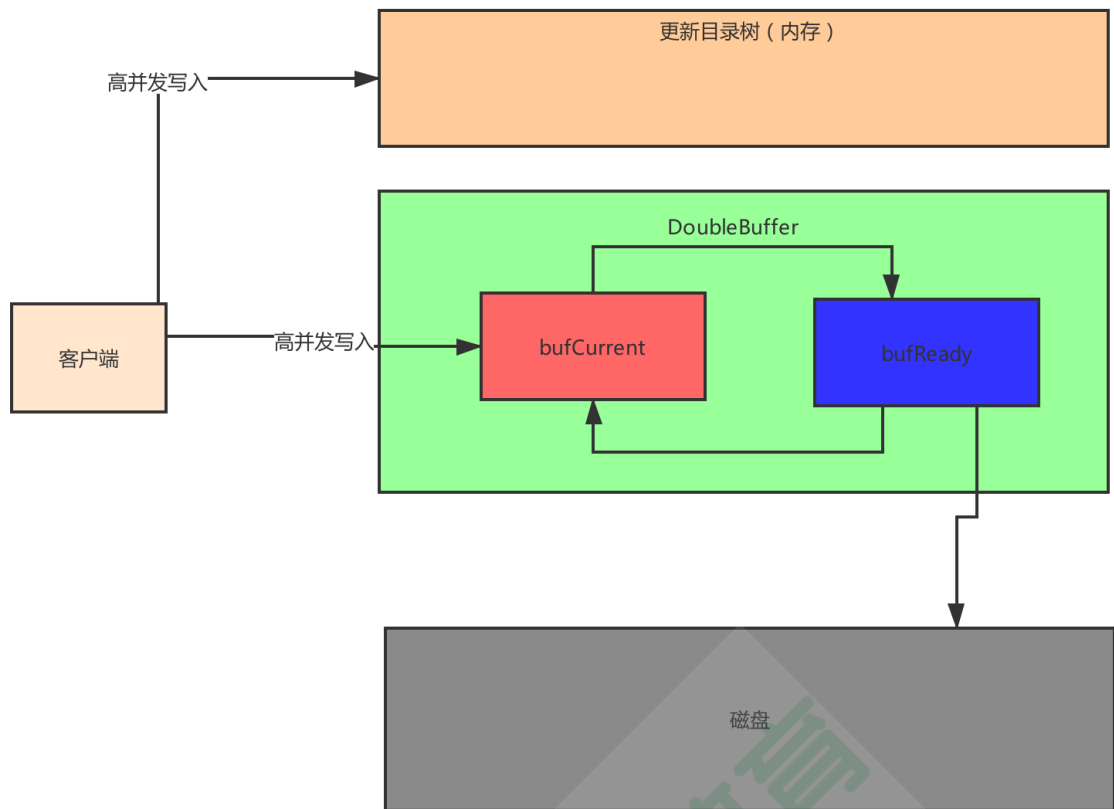


4.2 元数据管理（120分钟）

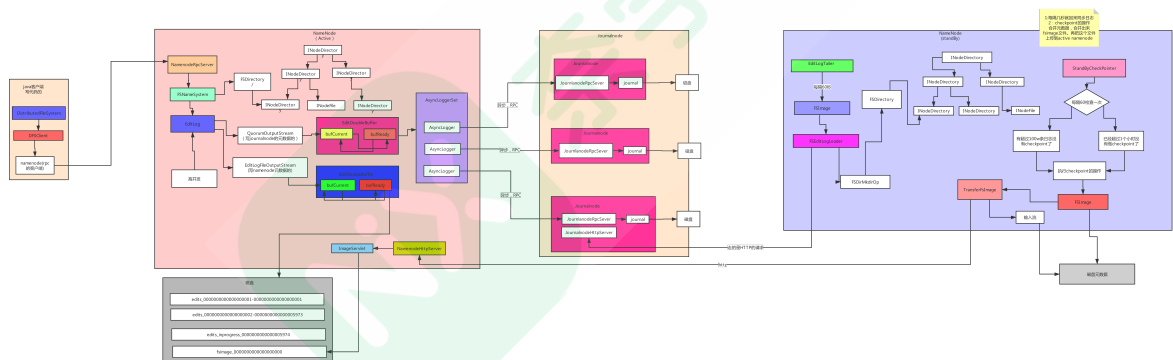
4.2.1 元数据管理流程回顾



4.2.2 元数据管理的双缓冲机制



4.2.3 元数据管理流程



五、优秀书籍推荐 (3分钟)

图书推荐：



六、总结（5分钟）

1. NameNode管理元数据
2. NameNode的双缓冲代码设计

七、作业

1. 跟着流程图把源码走一遍
2. 自己画一遍流程图
3. 敲一遍双缓冲方案的代码

八、互动

1. 提问学生是否能理解元数据管理的双缓冲机制

